Statistical Inference Project - Exponential Distributions and Central Theorem Limit

Conrad Wong

Overview

This project runs 1000 simulations of averages of 40 exponentials, with the objective of comparing the exponential distribution with the Central Theorem Limit

Simulations

Runs 1000 simulations of averages of 40 exponentials, with lambda = 0.2, and plots the distribution of the result

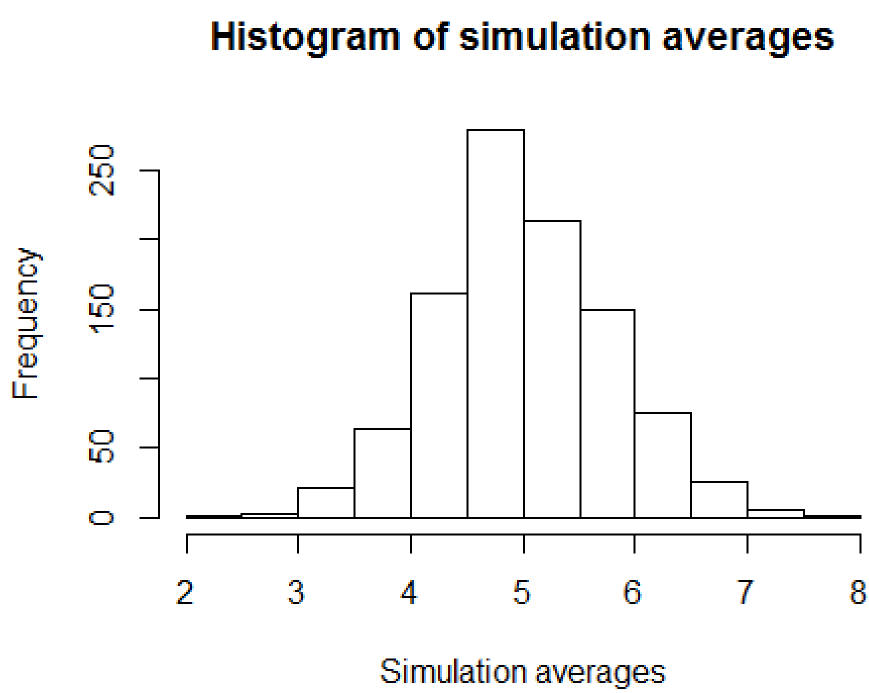
- sims is the number of simulations
- n is the number of averages

```
library (ggplot2)

set.seed(2398)

lambda <- 0.2
n <- 40
sims <- 1000

sim_means = NULL
for (i in 1:sims) sim_means = c(sim_means, mean(rexp(n, lambda)))
hist(sim_means, main="Histogram of simulation averages", xlab="Simulation averages")
```



1. Show the sample mean and compare it to the theoretical mean of the distribution.

Compares the sample mean to the theoretical mean (1/lambda):

```
sample_mean <- mean(sim_means)
theoretical_mean <- 1/lambda
```

Conclusion: The sample mean for the simulations (4.9960241) is very close to the theoretical mean (5).

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Compares the sample variance to the theoretical variance, calculated by: (1/lambda)^2 / n

```
sample_variance <- var(sim_means)
theoretical_variance <- 1 / (lambda)^2 / n
```

Conclusion: The variance of the sample (0.6119349) is very close to the theoretical variance (0.625)

3. Show that the distribution is approximately normal.

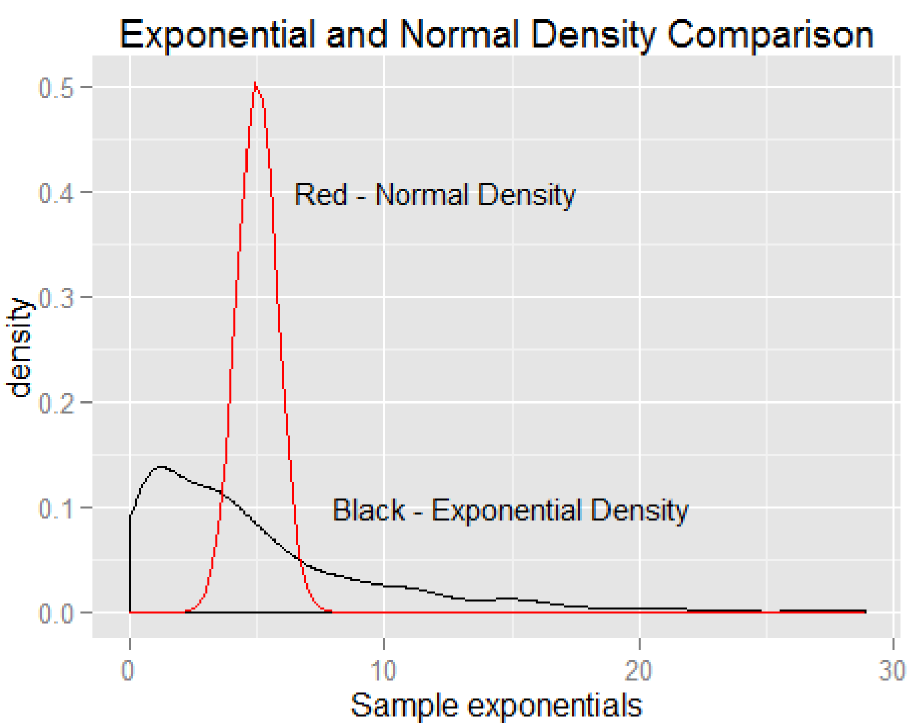
Generates 1000 samples from the exponential distribution and compares the density to the normal curve (using the theorical mean and standard deviation)

First, we'll compare the density of the exponential and normal distributions

```
exp <- rexp(sims, lambda)

g <- ggplot(data.frame(exp), aes(exp))
g <- g + geom_density()
g <- g + stat_function(fun=dnorm, arg=list(mean=5, sd=sqrt(0.625)), col="red")
g <- g + labs(title="Exponential and Normal Density Comparison", x="Sample exponentials")
g <- g + annotate("text", x = 12, y = 0.4, label = "Red - Normal Density", size=4)
g <- g + annotate("text", x = 15, y = 0.1, label = "Black - Exponential Density", size=4)


g
```



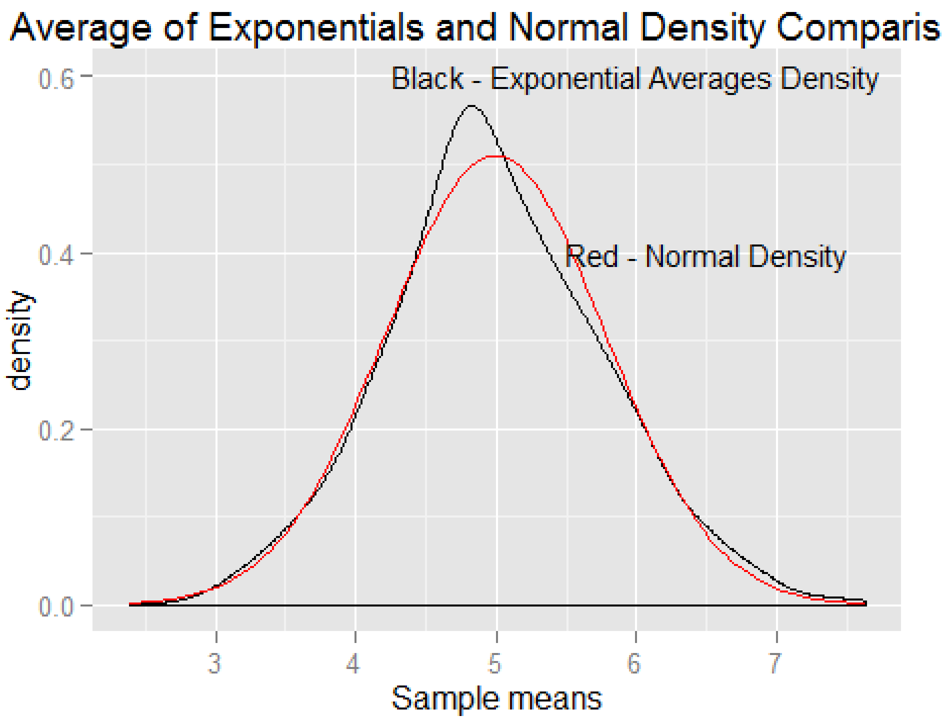
The densities of both distributions are very different

Now, let's compare the normal distribution to the density of the averages of exponentials we simulated.

```
g <- ggplot(data.frame(sim_means), aes(sim_means))
g <- g + geom_density()
g <- g + stat_function(fun=dnorm, arg=list(mean=sample_mean, sd=sqrt(sample_variance)), col="red")
```

```
g <- g + labs(title="Average of Exponentials and Normal Density Comparison", x="Sample means")
g <- g + annotate("text", x = 6.5, y = 0.4, label = "Red - Normal Density",  size=4)
g <- g + annotate("text", x = 6, y = 0.6, label = "Black - Exponential Averages Density", size=4)

g
```



Conclusion: The density of the exponential distribution is very different to the normal distribution. However, when you take the averages of 40 exponentials, the density becomes approximately normal, as stated by the Central Limit Theorem.